STATISTICAL MODELLING AND ANALYSIS OF TIME-SERIES DATA

1. Pre-enrolment Quiz

(1) (Differentiating with vectors and matrices.) Let $x = [x_1, \ldots, x_n]^T \in \mathbb{R}^n$, a column vector. For a function $x \in \mathbb{R}^n \to f(x) \in \mathbb{R}$, let

$$\nabla f(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix}.$$

Find ∇f for $f(x) = p^T x$ and $f(x) = x^T R x$ where p is a (column) vector and R is a square matrix.

(2) (Knowledge of the univariate Gaussian probability density function.) Let $p(x) = c \exp\left(ax^2 + bx\right)$ be a Gaussian probability density function, where x is a scalar. Find the its mean and variance in terms of the constants a, b and c.

(3) (Knowledge of the conditional probability density.) A bivariate Gaussian probability density function (pdf) is

$$f(x, y) = \frac{1}{2\pi \left|\Sigma\right|^{1/2}} \exp\left(-\frac{1}{2}\left[x - m_1, y - m_2\right] \Sigma^{-1} \left[x - m_1, y - m_2\right]^T\right),$$

where $\Sigma = \begin{bmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{bmatrix}$ is the called covariance matrix, and vector $(m_1, m_2)^T$ is called the mean vector.
Show the conditional pdf $f(x|y)$ is a Gaussian pdf with

$$\text{mean } = m_1 + \frac{\rho}{\sigma_2^2}(y - m_2) \qquad \text{and} \qquad \text{variance } = \sigma_1^2 - \frac{\rho^2}{\sigma_2^2}.$$

(Hint: write $f(x, y) = g(x, y)h(y)$ where the function $g(x, y)$ contains all the x and xy terms of $f(x, y)$ and then use the conditional pdf formula. Also, first attempt this question for $m_1 = m_2 = 0$.)

(4) (Basic knowledge of Monte Carlo.) Let X_i, $i = 1, \ldots, n$, be a collection of independent random variables with mean μ and variance σ^2. Let $\bar{X} = (X_1 + \ldots + X_n)/n$. Find $\mathbf{E}\left\{\left(X_1 - \bar{X}\right)^2\right\}$. Now find the constant c so that

$$c \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$$

is an unbiased estimate of σ^2.

S.S.Singh, University of Wollongong, June 2024

2. Solutions

(1) $\nabla f = x$ and $\nabla f = Rx + R^T x$.

(2) The Gaussian with mean m and variance σ^2 density is $\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-m)^2\right\}$. Expand the square and match terms with $c \exp\left(ax^2 + bx\right)$.

(3) To convey the idea, let $m_1 = m_2 = 0$. Expand the argument of the exponential

$$-\frac{1}{2\det}[x,y]\begin{bmatrix} \sigma_2^2 & -\rho \\ -\rho & \sigma_1^2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = -\frac{1}{2\det}\left(x^2\sigma_2^2 - 2yx\rho + \sigma_1^2 y^2\right)$$

$$\exp(\text{LHS}) = \exp\left(-\frac{1}{2\sigma_2^{-2}\det}\left(x^2 - \frac{2xy\rho}{\sigma_2^2} + s^2\right)\right) \times \exp\left(\frac{1}{2\sigma_2^{-2}\det}s^2 - \frac{1}{2\det}\sigma_1^2 y^2\right)$$

Let

$$g(x,y) = \exp\left(-\frac{1}{2\sigma_2^{-2}\det}\left(x^2 - \frac{2xy\rho}{\sigma_2^2} + s^2\right)\right).$$

We can write

$$f(x,y) = g(x,y)h(y)$$

where $g(x,y)$ contains all the x and xy terms of $f(x,y)$. The conditional pdf is

$$f(x|y) = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)\mathrm{d}x} = \frac{g(x,y)}{\int_{-\infty}^{\infty} g(x,y)\mathrm{d}x}.$$

There is no need to compute the integral in the denominator if we note that $g(x,y)$ is the exponential term of a Gaussian pdf if we choose

$$2sx = 2xy\frac{\rho}{\sigma_2^2}$$

or $s = y\rho/\sigma_2^2$. Thus $f(x|y)$ is a Gaussian density with mean $y\rho/\sigma_2^2$ and variance $\det/\sigma_2^2 = (\sigma_1^2\sigma_2^2 - \rho^2)/\sigma_2^2$.

(4) Show that $\mathbf{E}\left\{\left(X_1 - \bar{X}\right)^2\right\} = \sigma^2(n-1)/n$ and thus $c = 1/(n-1)$.